

Novel 18-gene signature for predicting relapse in ER-positive, HER2-negative breast cancer

Richard Buus^{1,2}, Belinda Yeo^{3,4}, Adam R Brentnall⁵, Marie Klintman⁶, Maggie Chon U Cheang⁷, Komel Khabra⁸, Ivana Sestak⁵, Qiong Gao¹, Jack Cuzick⁵, Mitch Dowsett^{1,2}

¹ The Breast Cancer Now Toby Robins Research Centre at the Institute of Cancer Research, London, UK

² Ralph Lauren Centre for Breast Cancer Research, Royal Marsden Hospital, London, UK

³ Olivia Newton-John Cancer Research Institute, Melbourne, Australia

⁴ Austin Health, Melbourne, Australia

⁵ Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, United Kingdom

⁶ Lund University, Skane University Hospital, Department of Clinical Sciences Lund, Oncology and Pathology, Lund, Sweden

⁷ Clinical Trials and Statistic Unit, The Institute of Cancer Research, London, UK

⁸ Research Data Management and Statistics Unit, Royal Marsden Hospital, London, UK

Corresponding Author:

Mitch Dowsett

Ralph Lauren Centre for Breast Cancer Research

Royal Marsden Hospital, Fulham Road

LONDON, SW3 6JJ, UK

Tel: +44 207 808 2885; Fax: +44 207 376 3918

E-mail: mitch.dowsett@icr.ac.uk

Disclosure of Potential Conflicts of Interest

Maggie Chon U. Cheang: Patents, Royalties, Other Intellectual Property: PAM50 patent (no financial gain in past 2 years)

Abstract

Purpose: Several prognostic signatures for early ER+ breast cancer have been established with a 10-year follow-up. We tested the hypothesis that signatures optimised for 0-5 year and 5-10 year follow-up separately are more prognostic than a single signature optimised for 10 years.

Experimental Design: Genes previously identified as prognostic or associated with endocrine resistance were tested *in silico* using Cox regression in 747 ER+/HER2- samples from postmenopausal patients treated with 5 years' endocrine therapy. RNA expression of genes found to be significant was assayed in primary ER+/HER2- tumours from 948 postmenopausal patients treated with 5 years' anastrozole or tamoxifen in the TransATAC cohort. Prognostic signatures for 0-10, 0-5 and 5-10 years were derived using Elastic Net modelling. Signature comparison was performed with Cox regression and likelihood ratio statistics. Validation was by a case:control (POLAR) study in 422 samples derived from a cohort of 1449.

Results: Ninety-three genes were selected by the *in silico* modelling; 63 of these were significantly prognostic in TransATAC, most similarly across each time period. Contrary to our hypothesis, the derived early and late signatures were not significantly more prognostic than the 10-year signature. The 10-year signature was validated in the POLAR case:control set and showed similar prognostic information to OncotypeDX, PAM-50 ROR, Breast Cancer Index and IHC4 in TransATAC.

Conclusions: The derived 10-year signature predicts risk of metastasis in patients with ER+/HER2- breast cancer similar to commercial signatures. The hypothesis that early and late prognostic signatures are significantly more informative than a single signature was rejected.

Introduction

Five years of adjuvant endocrine therapy is standard treatment for patients with primary oestrogen receptor positive (ER+) breast cancer, and clearly improves prognosis (1). Multiparameter molecular assays are increasingly used to estimate prognosis and guide treatment decisions of patients with primary ER+ breast cancer. These include the Oncotype DX Recurrence Score (RS) (2), Prosigna PAM50 (3), Breast Cancer Index (BCI) (4), EndoPredict (5) and IHC4 (6). All of them have been evaluated in the TransATAC series of samples that were established from patients with ER+ primary breast cancer randomised to treatment with 5 years' anastrozole or tamoxifen in the ATAC trial (7).

It has become clear that following surgery the risk of recurrence in ER+ primary breast cancer is not constant across all breast cancer subtypes, as evident between intrinsic subtypes. For example, data from 2985 tumours showed basal subtypes experiencing high recurrence rates early that decline after 5 years after surgery, contrary to patients categorised as luminal A, who experienced a lower but relatively constant recurrence rate during the first decade (8).

There are additional molecular differences beyond the subtype of a tumour. In TransATAC we have previously shown that the oestrogen-module of RS was prognostic within five years of surgery (during endocrine therapy), however it became non-informative for recurrences beyond five years thus weakening the overall prognostic value of RS (9). In the same data set, patients with high ER expression by RT-PCR were twice as more likely to have a relapse 5-10 years after surgery than within the first 5 years. Bianchini et.al. reported risk stratification by integrating the mitotic kinase score (MKS) and an ER-related score (ERS), both based on genes constituting the proliferation and oestrogen modules of RS. Women with high MKS and ERS tumours were at greater risk of late recurrence (10). More recently, improved risk estimation beyond 5 years by RS was reported when integrated with dichotomised ER expression assessed by RT-PCR (11).

Extending endocrine therapy beyond five years has been shown to reduce late recurrence rate (12, 13), however those most likely to benefit from such therapy need to be identified. While some of the widely used prognostic assays for ER+ patients have been shown to be prognostic for risk beyond five years (14-17), none of them have been optimised to quantify residual risk after five years free from recurrence and their ability to predict late relapse varies substantially (18). The different time-dependent performance of multiparameter molecular signatures indicates that molecular features of ER+ breast cancers may be identified to improve prediction of residual risk in order to spare those patients with significantly low risk of late recurrence from extended endocrine therapy.

We therefore hypothesised that prognostic signatures optimised specifically for the early (0-5 years) and late (5-10 years) follow-up periods respectively would be more prognostic than a single signature optimised for the whole 10-year follow-up period. To test this hypothesis we developed time-dependent prognostic signatures in patient samples from the TransATAC series for early, late and 10 year follow-up periods. The prognostic performance was tested in an independent sample set and against commercial signatures already assessed in TransATAC. Our primary aim was to compare the prognostic value of the newly developed signature(s) added to clinical treatment score (CTS) (6) and that of PAM50 ROR-P added to CTS.

Materials and methods

Patient cohorts

An *in silico* analysis drew on four published breast cancer cohorts (GSE6532, GSE9195, GSE17705, GSE26971) analysed on either Affymetrix HG-U133A (GPL96) or HG-U133 Plus 2.0 (GPL570) microarray platforms. The two platforms shared 22,277 probes to which we restricted our analyses. This cohort had 747 unique patient samples that matched our selection criteria: ER+, HER2-, treated with five years of endocrine therapy, chemotherapy naive, with either distant metastasis-free survival (DMFS) or relapse-free survival (RFS) available with a long follow-up. Details of the inclusion criteria are listed in Supplementary Methods and a full list of samples included in the analysis is shown in Supplementary Table S1.

In the TransATAC cohort, RNA was available from 948 formalin-fixed paraffin-embedded (FFPE) tumours from the ATAC (Arimidex, Tamoxifen, Alone or in Combination) trial, previously extracted by Genomic Health Inc. (GHI) (19). Eligibility required hormone receptor (HR)+, HER2- disease, without chemotherapy treatment and at least 500ng RNA available. One hundred and eighty three recurrence events were recorded for this cohort. This study was approved by the South-East London Research Ethics Committee and all patients gave informed consent.

The POLAR (Predictors Of early versus Late Recurrence in ER+ breast cancer) samples were identified from archives of The Royal Marsden Hospital (RMH), London, United Kingdom and Lund University Hospital (LUH) Biobank, Sweden. Eligibility criteria were patient with ER+, HER2- early breast cancer diagnosed between January 2000 and December 2004, treated with curative intent and with a follow-up data cut-off at May 2014. Patients must have received five years of adjuvant endocrine therapy (unless relapse occurred within this time); (neo)adjuvant chemotherapy was permitted. A 422 sample case:control design was used; controls were randomly selected according to matching criteria from the remaining cohort of patients who did not relapse during follow-up. The total number of patients that was drawn upon was 1449. The four matching criteria used in this study were: 1) age at diagnosis (<50 or >50 years), 2) NPI category (<3.4; 3.4-5.4, >5.4), 3) type of adjuvant endocrine therapy (tamoxifen only vs. any aromatase inhibitor (AI)) and 4) chemotherapy use (yes or no). Two-hundred and forty-seven recurrence events were recorded. The POLAR study was approved by the RMH Research Ethics Committee (CCR: 4122) and the ethics committee of Lund University Hospital (LU 240-01).

Study end points

The primary endpoint was time to any recurrence that was defined as either locoregional (ipsilateral breast, contralateral breast and regional lymph nodes) and/or distant recurrence. Secondary endpoint was time to distant recurrence, which was the time from diagnosis until metastasis from the primary tumour at distant organs, excluding contralateral disease and locoregional and ipsilateral recurrences. Death before recurrence was treated as a censoring event for both endpoints.

Analytic procedures

In the *in silico* data set 454 probes representing 454 genes (Supplementary table 3) were analysed at univariate level; those significant in univariate analyses in a particular setting were entered into multivariate analyses. Additional details are in the Supplementary Methods.

For TransATAC, RNA was extracted by GHI for the RS study (19). One-hundred ng RNA was used with the nCounter platform (NanoString Technologies, Seattle, WA) to assay the 93 endogenous and 7 reference genes selected in the process of the *in silico* analysis in 948 TransATAC samples.

For POLAR, RNA was extracted from three 3x10micron unstained sections with more than 40% tumour cellularity using RNeasy FFPE kit (Qiagen) according to the manufacturer's instructions. RNA was quantified by Nanodrop (Thermo Fisher Scientific). Between 50 and 200ng RNA was used to profile the expression of 27 endogenous and 5 reference genes with the NanoString nCounter.

NanoString expression data was background corrected by subtracting the mean of the eight negative control probes, normalised with the geometric mean of five reference genes that had a correlation of Pearson's $r > 0.8$ with all endogenous genes. The data set was then log (base 2)-transformed and z-score transformed. KIF20A gene was detected in <10% of samples in the TransATAC cohort and was removed from the data set. CTS, which carries information on tumour size, nodal status, grade, age and type of endocrine therapy was calculated as published previously (6).

We trained separate early, late and 10-year signatures by performing elastic net analysis in the TransATAC training cohort. Our objective was to test if the early and late signatures had statistically significantly more prognostic power than the 10-year signature. If so, we would test the validity of the early and late signatures in the non-chemotherapy-treated subpopulation of POLAR and also test their performance in the chemotherapy-treated POLAR cohort. If the early and late signatures were not statistically significantly more prognostic than the overall signature, we would test the validity of the overall signature in the chemotherapy-naïve POLAR group and explore its performance in the chemotherapy-treated POLAR group.

Statistical analyses in the *in silico* cohort were carried out at the ICR using R version 3.03. Statistical analyses using the TransATAC cohort were performed at Queen Mary University of London with STATA version 13.1 and R version 3.0.3. Statistical work on POLAR was carried out at the Royal Marsden Hospital using the Statistical Analysis Plan version 2.0 and PRISM 6.0c. The statistical analysis plan for the TransATAC study was approved by the Long-term Anastrozole vs Tamoxifen Treatment Effects (LATTE) committee and for the POLAR study by the RMH Committee for Clinical Research before data analysis took place and is described in the Supplementary Methods. All statistical tests were two-sided.

Results

Candidate gene selection and *in silico* analysis

In order to derive time-dependent prognostic signatures we shortlisted 585 candidate genes representing proliferation, oestrogen signalling, immune infiltration and immune signalling. These genes were tested for prognostic significance *in silico* from publically available gene expression sets in ER+ endocrine treated breast cancer. A flow-chart illustrating the approach is shown in Figure 1.

Sixty-seven genes of interest that are part of the PAM50, OncotypeDX RS, EndoPredict and BCI profilers were also included. Additional genes likely to be related to benefit from endocrine therapy were identified from 81 patients by reanalysing our previously published neoadjuvant endocrine therapy treated set of samples (20) (<https://www.synapse.org/#!Synapse:syn16243>). From this dataset we identified 164 candidate genes by examining correlation of individual gene expression from untreated biopsies with change in: 1) Ki67, 2) proliferation-associated gene cluster, 3) oestrogen-associated gene cluster, 4) expression of mGIDE (modified version of the Global Index of Dependence on Estrogen) (21) genes following two weeks of AI treatment. An additional 354 genes were selected based on literature searches. Genes from published gene modules of the proliferation-associated gene cluster, oestrogen-associated gene cluster and inflammatory response signature (20), the tumour invasion/metastasis module (PLAU) (22) and IGG-14 module (immunoglobulin gamma) (23) were also included. The complete list of candidate genes and the reason for their inclusion is detailed in Supplementary Table 1.

Seven hundred and forty-seven samples from microarray expression dataset was compiled from four publically available breast cancer cohorts to investigate the relationship between genes and outcome (Supplementary table 2) (5, 24-26). Expression data was available for 454 genes (Supplementary table 3). We performed univariate Cox proportional hazards regression analyses for early, late and 10-year follow-up periods using relapse-free survival (RFS) and distant metastasis-free survival (DMFS) as end-points respectively that identified 212 genes that were significant at $p < 0.01$ in any of the analyses (Supplementary Table 4). Genes significantly prognostic in a particular time-period were taken forward for multivariate analyses performed by Cox proportional hazards regression with DMFS and RFS as endpoints respectively in the early, late and 10-year follow-up settings. This resulted in 88 genes being selected in the models (Supplementary Table 5), of which 17 genes were removed due to high correlation of expression with other candidates already selected (Supplementary Table 6). An additional 29 genes were added that included candidates without probes available in the *in silico* analyses, some recently emerging candidates and also seven reference genes (Supplementary Table 7).

Expression profiling and signature building in TransATAC

Sample availability in TransATAC is shown in Figure 2a. Expression data for the 100 selected genes (including housekeeping genes; Supplementary Table 9) were obtained for 948 patient samples in TransATAC using the NanoString nCounter. We assessed the prognostic value of these molecular variables in TransATAC for early, late and 10-year time periods for RFS. Sixty-three genes were statistically significant in at least one of the time-windows assessed (Supplementary Figure 1, Supplementary Table 7). We found different prognostic properties between early and late periods for 20 genes. Six genes were prognostic early but not in the late period (CD79, IL6ST, LRRC48, MPZL1, PGR, PIGV); 14 genes were not significantly prognostic early but gained prognostic significance in the late setting (ANP32E, ANXA1, CTSL2, EPB41L2, ESR1, FOXA1, ICOS, IL17RB, MMP9, MYCBP2, NR2F1, PDZK1, SLAMF8 and TCF7L2).

The TransATAC cohort was then randomly split into 2/3 ($n=634$) training and 1/3 ($n=314$) validation sets while ensuring that the recurrence rate was similar in the two subgroups. Demographics for the training, validation and overall cohorts are presented in Supplementary Table 10. We aimed to select prognostic variables independent of clinico-pathological features that are commonly used for

prognosis. To achieve this, on top of the 63 statistically significant genes in univariate analyses, CTS was also entered into multivariate selections for early, late and 10-year time-periods respectively. Elastic Net Penalised Cox regression with leave one out cross-validation was used for feature selection in the TransATAC training set. CTS was selected in all three signatures in addition to 18 genes in the 10-year, 16 genes in the early and 15 genes in the late follow-up analyses. The variables and their coefficients derived from the elastic net models are listed in Table 1. CTS had the highest coefficient in each of the time periods.

Comparison of time period-optimised prognostic signatures in TransATAC validation set

TransATAC was used to validate and compare the prognostic information of the three time period-dependent signatures (Table 2). In the 0-10 years follow-up period all three newly derived signatures were significantly prognostic, with the late signature being significantly less informative than the 10-year signature (10-year signature $LR\chi^2=28.0$; early signature $LR\chi^2=33.4$; late signature $LR\chi^2=18.1$). In the 0-5 years period the 10-year signature and early signature were equally prognostic and significantly more than the late signature ($LR\chi^2$: 10-year signature=14.1; $LR\chi^2$: early signature=14.9; $LR\chi^2$: late signature=8.9). In the late setting the early signature was the most prognostic, followed by the 10-year and late signatures ($LR\chi^2$: 10-year signature=13.9; $LR\chi^2$: early signature=18.6; $LR\chi^2$: late signature = 9.3). CTS was strongly prognostic in all three time periods (CTS 0-10y $LR\chi^2=48.7$; CTS 0-5y $LR\chi^2=29$; CTS 5-10y $LR\chi^2=19.8$).

For the 0-10 year period all three signatures added statistically significant prognostic information beyond that of the CTS ($\Delta LR\chi^2$: 10-year signature=7.9; $\Delta LR\chi^2$: early signature=10.3; $\Delta LR\chi^2$: late signature=4.3). In the 0-5 years period none of the signatures added significant prognostic information to CTS. However in the 5-10 year period the 10-year and early signatures added statistically significant prognostic information to CTS (10-year signature $\Delta LR\chi^2=4.8$; early signature $\Delta LR\chi^2=8.0$; late signature $\Delta LR\chi^2=2.7$).

Given that the early- and the late-signatures were not statistically significantly more prognostic than the 10-year signature in the respective periods they were optimised for, we rejected our primary hypothesis that signatures optimised separately for the early and the late follow-up periods respectively are more prognostic than a 10-year signature but we proceeded to assess the validity of the 18-gene, 10-year signature in an independent cohort and to compare its performance with that of commercial signatures.

Signature test of 10-year validity in POLAR cohort

A matched case-control set of samples was compiled from the Royal Marsden Hospital and Lund University Hospital archives (POLAR) to validate the 10-year signature (Figure 2b, Supplementary Table 11). Our aims were to test the validity the 10-year signature in an endocrine-therapy only cohort similar to the training set, and also to explore if the prognostic property (if any) extends to a higher risk, chemotherapy-treated population. The latter cohort was of interest in the 5-10 year period because of the potential for its use in selecting patients for extended adjuvant endocrine therapy.

Despite having matched cases and controls on NPI category, the CTS was still higher in cases than in controls: 201.9+/-98 (SD) vs. 170.8+/-87.6 ($p=0.0009$) respectively. In a univariate analysis, CTS had an Odds Ratio (OR) of 1.004 (95% CI 1.001-1.006) for a one unit increase. We assessed a multivariate model with CTS with and the 10-year signature; both were found to be statistically significant: 10-year signature OR=1.851 (95%CI 1.194-2.868), $p=0.006$; CTS OR=1.003 (1.001-1.005), $p=0.012$.

We also assessed whether the 10-year signature added significant prognostic information above CTS alone using likelihood ratio tests (Table 3). In the overall POLAR cohort ($n=422$), CTS was prognostic across 10 years and in the early follow-up period (CTS 0-10 years period $LR\chi^2=11.23$; 0-5 years period $LR\chi^2=22.09$) but not in the 5-10 year period. The 10-year signature was prognostic in all three follow-up periods and contributed to CTS with significant prognostic information in the 10-year and early periods (0-10 years period $\Delta LR\chi^2$: CTS+10-year signature vs. CTS=7.74; 0-5 years period $\Delta LR\chi^2$: CTS+10-year signature vs. CTS=7.59) but not in the 5-10 year period. Both CTS and the 10-year signature were marginally more informative across the 10 years in the chemotherapy-treated POLAR cohort compared with the endocrine only population despite the latter having more patients and events (patients: $n=170$ vs. $n=252$; events: 99 vs. 148). Additionally, the 10-year signature added significantly more prognostic information to CTS in the chemotherapy-treated group ($\Delta LR\chi^2$: CTS+10-year signature vs. CTS=6.71) when compared to those receiving endocrine therapy only ($\Delta LR\chi^2$: CTS+10-year signature vs. CTS=2.47).

Prognostic properties of the 18 individual genes constituting the 10-year signature were assessed in POLAR and compared to data obtained in TransATAC. In POLAR only 8 out of the 18 genes were significantly prognostic at the univariate level (Supplementary Figure 2) but all genes except tumour necrosis factor alpha (TNF) showed the same prognostic direction both in TransATAC and POLAR.

Comparison of the 10-year signature with CTS, RS, PAM50 ROR, BCI and IHC4 in TransATAC

We have previously published data on the prognostic performance of CTS, RS, PAM50 ROR, BCI and IHC4 in TransATAC (6, 16, 19, 27); data for all scores was available for 271 patients in the validation cohort. We assessed their prognostic information for 10 years after surgery using any recurrence and distant recurrence as endpoints respectively and compared them with the newly developed 10-year signature (Table 4). For both any and distant recurrence the BCI provided the most added information beyond the CTS in this set (any recurrence: CTS: $LR\chi^2=37.4$; BCI: $\Delta LR\chi^2=9.5$; distant recurrence: CTS: $LR\chi^2=46.7$; BCI: $\Delta LR\chi^2=14.5$, respectively). The novel 10-year signature performed similarly to the other three scores in this respect.

Discussion

We developed novel time-specific prognostic signatures for early, late and 10-year follow-up periods for ER+, HER2- patients treated with endocrine therapy alone to allow us to test the hypothesis that sequentially applying early and late signatures could be more prognostic for risk of relapse than a single newly developed 10-year signature. This hypothesis was largely based around our observation that the performance of some components in many of the commercially available signatures varied between these time periods. For example, we found that ESR1 and the oestrogen module overall in the RS was less prognostic in years 5-10 than 0-5 (9). Analogous findings were made by Bianchini et

al (10). Very recently the EBCTCG published data on clinicopathological and limited immunohistochemical data on over 60,000 women that were treated with 5 years of endocrine therapy (28). While progesterone receptor showed strong prognostic performance in years 0-5, it showed no significant relationship with prognosis thereafter. These data on markers associated with hormone responsiveness support the contention, but by no means prove, that cessation of endocrine treatment at 5 years may lead to increased recurrence risk in more hormonally responsive tumours. We therefore included in our assessment genes that we and others have found to be associated with the antiproliferative response of primary ER+ breast cancer to oestrogen deprivation. Our work involved an *in silico* discovery set of 747 samples, training and test sets of 634 and 314 TransATAC samples, respectively, and independent case:control series from 1449 eligible samples. As such this was one of the largest original gene expression analyses undertaken for evaluating prognosis in ER+ breast cancer.

Of the 92 genes selected from *in silico* data and assessed in univariate analyses in TransATAC we found 63 to be significantly prognostic ($P < 0.05$) in any of the three time periods which is considerably more than expected by chance after allowing for multiple testing errors. For most genes the same prognostic pattern was observed for early and late periods however we observed some possibly different prognostic properties for 20 genes. Notably, consistent with the above arguments, higher levels of ESR1 and its pioneer factor FOXA1 showed a shift at 5 years to be associated with worse prognosis beyond 5 years but surprisingly over the 10 year period the two genes were associated with poor prognosis. The complementary role whereby upon stimulus ER binding to chromatin is dependent on the presence of FOXA1 is well established (29). In our dataset FOXA1 and ESR1 correlated highly (Pearson's $R: 0.65$); the possibility that increased expression of one or both may put patients at increased risk of late relapse merits further investigation, particularly with regard to whether the genes also identify patients who benefit from extended adjuvant therapy.

The optimised time-dependent signatures derived in the TransATAC training set were rather similar to one another in makeup. All genes in the 10-year signature featured in either (or both) of the early and late signatures with their coefficients being in the same direction. The early and late signatures had five and three variables respectively not present in the 10-year signature suggesting that the early and late signatures may not have captured time-specific features or that such time-specific features that exist exert a minor modulatory influence on the overall prognosis over 10 years. It is notable that CTS was consistently the most prognostic variable in the three time-dependent models and that its contribution was similar in both early and late recurrence. This is consistent with the data of the EBCTCG that classical clinico-pathologic features retain their strong prognostic influence beyond 5 years (28).

Given that the 10-year signature captured prognostic features of both early and late events, it is perhaps not surprising that no improvement was seen in the use of early and late signatures compared to the overall 10-year signature that led to the rejection of our hypothesis. Also, it should be noted that splitting of the 0-10 year time period into 0-5 and 5-10 year periods markedly reduces the power to detect prognostic contributions. At least a contributory factor for the lack of improvement may be because of the dominance of proliferation-related genes in our and other signatures. As shown in our earlier analysis of the RS, each of the individual proliferation genes and the integrated module are equally prognostic before and after 5 years (9). Notably this is also

supported by the observation by the EBCTCG that Ki67 was equally prognostic before and after 5 years in their overview analysis of late recurrence (28).

The 10-year signature was nonetheless validated in the POLAR sample set and provided significant prognostic information in both chemotherapy-naïve and -treated cohorts. Moreover, it added independent prognostic information beyond that of CTS in the POLAR cohort. Comparison of the information provided by each gene showed that eight out of the 18 genes were significantly prognostic at univariate level in POLAR (four genes at $P < 0.05$, two genes at $P < 0.01$ and three genes $P < 0.001$). TNF showed opposite prognostic direction in training and validation sets, thus weakening the performance of the signature in POLAR. TNF is a versatile pro-inflammatory cytokine that has both pro- and anti-tumour activities promoting lymphocytic infiltration, activating the nuclear factor- κ B, c-Jun N-terminal kinase and mitogen-activated protein kinase pathways and is capable of inducing apoptosis through TNF receptor 1 and 2 (30). It may be that the inclusion of higher risk, chemotherapy-treated patients in POLAR contributed to the difference in TNF's prognostic pattern; further investigation is needed to explain the relationship of TNF and risk of relapse in these cohorts.

The 10-year signature was compared to established prognostic signatures in the TransATAC validation set. Importantly, the 10-year signature was developed for the endpoint of any recurrence contrary to the endpoint of distant recurrence used in the development of RS, PAM50 ROR, BCI and IHC4. In univariate assessments, BCI and the 10-year signatures were the most informative for both all and distant recurrence. When added to CTS, all signatures assessed provided similar amount of information, with CTS+BCI being the most informative for distant recurrence. This new signature did not outperform the established signatures even though it was based on a large and wide-ranging analysis of both established prognostic genes and novel genes with a clear rationale for inclusion. It seems unlikely that a step-change in prognostic performance with any further elaborations on gene expression profiles in this context; other approaches that assess response to treatment or integrate mutational profiles or by the use of circulating tumour DNA (ctDNA) are likely to be more fruitful.

The results presented here support the mounting evidence that better risk estimation can be achieved by combining molecular profilers with clinico-pathological factors. For the three time-dependent signatures derived in TransATAC CTS was the most prognostic in all three time-dependent signatures and provided more prognostic information than RS, ROR, BCI and IHC4 respectively. Additionally, all profilers added significant prognostic information to CTS leading to combined signatures being significantly more informative.

Our study has strengths and limitations. An advantage was that a large discovery cohort of 634 samples was used for signature training. All tumours were ER+, HER2- from postmenopausal patients who had five years of endocrine therapy without chemotherapy. This was a homogeneous group of breast cancers which reduced confounding factors such as tumour subtype, differing treatment lengths and types. Data for the clinical prognostic tests were obtained by the same methods as set out by the tests' developers. The same batch of RNA was used for the newly developed signatures presented here and for the clinical prognostic tests used in the comparisons, reducing intra-sample variation. The clinical data were derived from a registration standard trial with comprehensive follow-up over 10 years. Limitations include that CTS, IHC4 and the 10-year signature were derived in TransATAC therefore their performance in the comparisons was slightly overestimated compared to what we would see in independent cohorts. Also, although this study was relatively large compared

to others, the splitting of the data into early and late signatures decreased the statistical power for comparisons within those time periods.

In summary, we found that early and late signatures are unlikely to be more informative for predicting relapse than a single signature optimised for 10 years. Further development of gene expression signatures for prognosis in endocrine-treated ER+ breast cancer patients is unlikely to achieve a substantial improvement in performance.

Acknowledgements

We would like to acknowledge professor Mårten Fernö and Professor Per Malmström for their work on providing data and samples for the Lund cohort. We thank Genomic Health Inc., NanoString Technologies and BioTheranostics for the data of their respective gene signatures.

Funding

This work was supported by Breast Cancer Now, working in partnership with Walk the Walk and by the NIHR RM/ICR Biomedical Research Centre. ARB was funded by Cancer Research UK (Grant number C569/A16891). The study was supported by funds from Skåne County Council's Research and Development Foundation, Governmental Funding of Clinical Research within the National Health Service, grant numbers ALFSKANE-350191 (MK), the Swedish Breast Cancer Association (BRO), the Mrs Berta Kamprad Foundation and The Inger Persson Research Foundation. IS and JC were supported by Cancer Research UK (Programme Grant C569/A10404).

List of Tables and Figures

Table 1. Variables and corresponding beta-coefficients of the time-dependent 10-year, early and late signatures.

Table 2. HRs (95% CI) per standard deviation, $LR\chi^2$, p-values and c-indices for CTS, 10-year, early and late signatures in TransATAC validation cohort.

Table 3. $LR\chi^2$ and p-values for CTS and 10-year signature in three groups of POLAR validation set.

Table 4. HRs (95% CI) per standard deviation, $LR\chi^2$, p-values and c-indices for CTS, 10-year signature, BCI, RS, ROR and IHC4 for all and distant recurrences respectively in the TransATAC validation cohort.

ATAC = Arimidex, Tamoxifen, Alone or in Combination; ER = oestrogen receptor; PgR = progesterone receptor; RMH = Royal Marsden Hospital; LUH = Lund University Hospital; ET = endocrine Therapy.

Supplementary Table S1. List of 585 candidate genes

Supplementary Table S2. List and identifiers for the 747 patient in silico cohort

Supplementary Table S3. List of 454 Affymetrix probes studied in silico

Supplementary Table S4. List of 212 genes significantly prognostic ($p < 0.01$) in any of the three time periods studied in silico

Supplementary Table S5. List of 88 genes by multivariate selections in any of the three time periods studied in silico

Supplementary Table S6. List of 17 genes manually removed from the multivariate list in silico

Supplementary Table S7. List of 29 genes added to the candidate list

Supplementary Table S8. Details of the 100 probe NanoString code set used in TransATAC

Supplementary Table S9. Hazard ratios, confidence intervals and p-values for the 92 genes assessed in TransATAC in univariate analyses

Supplementary Table S10. Demographics of TransATAC cohort

Supplementary Table S11. Demographics of POLAR cohort

IDC = invasive ductal carcinoma; ILC = invasive lobular carcinoma; PgR = progesterone receptor; NPI = Nottingham Prognostic Index; AI = aromatase inhibitor; RMH = LUH = ; POLAR = molecular Predictors Of early versus LAte Recurrence in ER+ breast cancer. * denotes matching criteria.

Figure 1. Flowchart of gene signature derivation *in silico* and in TransATAC cohort.

QC = Quality control; ATAC= Arimidex, Tamoxifen, Alone or in Combination.

Figure 2. CONSORT diagram of the availability of samples for analysis from the (a) ATAC trial and (b) POLAR collection of samples.

Supplementary Figure 1. Forest plot of hazard ratios and confidence intervals for the 92 genes assessed in TransATAC in univariate analyses. * denotes significance.

Supplementary Figure 2. Hazard ratios and odds ratios for the 10-year signature genes in TransATAC and POLAR respectively

References

1. Dowsett M, Forbes JF, Bradley R, Ingle J, Aihara T, Bliss J, et al. Aromatase inhibitors versus tamoxifen in early breast cancer: patient-level meta-analysis of the randomised trials. *Lancet* 2015;386:1341-52.
2. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004;351:2817-26.
3. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* 2009;27:1160-7.

4. Ma XJ, Salunga R, Dahiya S, Wang W, Carney E, Durbecq V, et al. A five-gene molecular grade index and HOXB13:IL17BR are complementary prognostic factors in early stage breast cancer. *Clin Cancer Res* 2008;14:2601-8.
5. Filipits M, Rudas M, Jakesz R, Dubsy P, Fitzal F, Singer CF, et al. A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. *Clin Cancer Res* 2011;17:6012-20.
6. Cuzick J, Dowsett M, Pineda S, Wale C, Salter J, Quinn E, et al. Prognostic value of a combined estrogen receptor, progesterone receptor, Ki-67, and human epidermal growth factor receptor 2 immunohistochemical score and comparison with the Genomic Health recurrence score in early breast cancer. *J Clin Oncol* 2011;29:4273-8.
7. Forbes JF, Cuzick J, Buzdar A, Howell A, Tobias JS, Baum M. Effect of anastrozole and tamoxifen as adjuvant treatment for early-stage breast cancer: 100-month analysis of the ATAC trial. *Lancet Oncol* 2008;9:45-53.
8. Voduc KD, Cheang MC, Tyldesley S, Gelmon K, Nielsen TO, Kennecke H. Breast cancer subtypes and the risk of local and regional relapse. *J Clin Oncol* 2010;28:1684-91.
9. Dowsett M, Sestak I, Buus R, Lopez-Knowles E, Mallon E, Howell A, et al. Estrogen Receptor Expression in 21-Gene Recurrence Score Predicts Increased Late Recurrence for Estrogen-Positive/HER2-Negative Breast Cancer. *Clin Cancer Res* 2015;21:2763-70.
10. Bianchini G, Pusztai L, Karn T, Iwamoto T, Rody A, Kelly C, et al. Proliferation and estrogen signaling can distinguish patients at risk for early versus late relapse among estrogen receptor positive breast cancers. *Breast Cancer Res* 2013;15:R86.
11. Wolmark N, Mamounas EP, Baehner FL, Butler SM, Tang G, Jamshidian F, et al. Prognostic Impact of the Combination of Recurrence Score and Quantitative Estrogen Receptor Expression (ESR1) on Predicting Late Distant Recurrence Risk in Estrogen Receptor-Positive Breast Cancer After 5 Years of Tamoxifen: Results From NRG Oncology/National Surgical Adjuvant Breast and Bowel Project B-28 and B-14. *J Clin Oncol* 2016;34:2350-8.
12. Goss PE, Ingle JN, Martino S, Robert NJ, Muss HB, Piccart MJ, et al. A randomized trial of letrozole in postmenopausal women after five years of tamoxifen therapy for early-stage breast cancer. *N Engl J Med* 2003;349:1793-802.
13. Davies C, Pan H, Godwin J, Gray R, Arriagada R, Raina V, et al. Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial. *Lancet* 2013;381:805-16.
14. Dubsy P, Brase JC, Jakesz R, Rudas M, Singer CF, Greil R, et al. The EndoPredict score provides prognostic information on late distant metastases in ER+/HER2- breast cancer patients. *Br J Cancer* 2013;109:2959-64.
15. Sestak I, Dowsett M, Zabaglo L, Lopez-Knowles E, Ferree S, Cowens JW, et al. Factors predicting late recurrence for estrogen receptor-positive breast cancer. *J Natl Cancer Inst* 2013;105:1504-11.
16. Sgroi DC, Sestak I, Cuzick J, Zhang Y, Schnabel CA, Schroeder B, et al. Prediction of late distant recurrence in patients with oestrogen-receptor-positive breast cancer: a prospective comparison of the breast-cancer index (BCI) assay, 21-gene recurrence score, and IHC4 in the TransATAC study population. *Lancet Oncol* 2013;14:1067-76.
17. Sestak I, Cuzick J, Dowsett M, Lopez-Knowles E, Filipits M, Dubsy P, et al. Prediction of late distant recurrence after 5 years of endocrine treatment: a combined analysis of patients from the Austrian breast and colorectal cancer study group 8 and arimidex, tamoxifen alone or in combination randomized trials using the PAM50 risk of recurrence score. *J Clin Oncol* 2015;33:916-22.
18. Zhao X, Rodland EA, Sorlie T, Vollan HK, Russnes HG, Kristensen VN, et al. Systematic assessment of prognostic gene signatures for breast cancer shows distinct influence of time and ER status. *BMC Cancer* 2014;14:211.

19. Dowsett M, Cuzick J, Wale C, Forbes J, Mallon EA, Salter J, et al. Prediction of risk of distant recurrence using the 21-gene recurrence score in node-negative and node-positive postmenopausal patients with breast cancer treated with anastrozole or tamoxifen: a TransATAC study. *J Clin Oncol* 2010;28:1829-34.
20. Dunbier AK, Ghazoui Z, Anderson H, Salter J, Nerurkar A, Osin P, et al. Molecular profiling of aromatase inhibitor-treated postmenopausal breast tumors identifies immune-related correlates of resistance. *Clin Cancer Res* 2013;19:2775-86.
21. Mackay A, Urruticoechea A, Dixon JM, Dexter T, Fenwick K, Ashworth A, et al. Molecular response to aromatase inhibitor treatment in primary breast cancer. *Breast Cancer Res* 2007;9:R37.
22. Desmedt C, Haibe-Kains B, Wirapati P, Buyse M, Larsimont D, Bontempi G, et al. Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. *Clin Cancer Res* 2008;14:5158-65.
23. Fan C, Prat A, Parker JS, Liu Y, Carey LA, Troester MA, et al. Building prognostic models for breast cancer patients using clinical variables and hundreds of gene expression signatures. *BMC Med Genomics* 2011;4:3.
24. Loi S, Haibe-Kains B, Desmedt C, Lallemant F, Tutt AM, Gillet C, et al. Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *J Clin Oncol* 2007;25:1239-46.
25. Loi S, Haibe-Kains B, Desmedt C, Wirapati P, Lallemant F, Tutt AM, et al. Predicting prognosis using molecular profiling in estrogen receptor-positive breast cancer treated with tamoxifen. *BMC Genomics* 2008;9:239.
26. Symmans WF, Hatzis C, Sotiriou C, Andre F, Peintinger F, Regitnig P, et al. Genomic index of sensitivity to endocrine therapy for breast cancer. *J Clin Oncol* 2010;28:4111-9.
27. Dowsett M, Sestak I, Lopez-Knowles E, Sidhu K, Dunbier AK, Cowens JW, et al. Comparison of PAM50 risk of recurrence score with oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy. *J Clin Oncol* 2013;31:2783-90.
28. Hongchao Pan RG, Jeremy Braybrooke, Christina Davies, Carolyn Taylor, Paul McGale, Richard Peto, Kathleen I. Pritchard, Jonas Bergh, Mitch Dowsett, and Daniel F. Hayes. Early breast cancer: 20-year risks if endocrine treatment stops at 5 years. *N Engl J Med* in press.
29. Hurtado A, Holmes KA, Ross-Innes CS, Schmidt D, Carroll JS. FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nat Genet* 2011;43:27-33.
30. Wang X, Lin Y. Tumor necrosis factor and cancer, buddies or foes? *Acta Pharmacol Sin* 2008;29:1275-88.

Figures and Tables

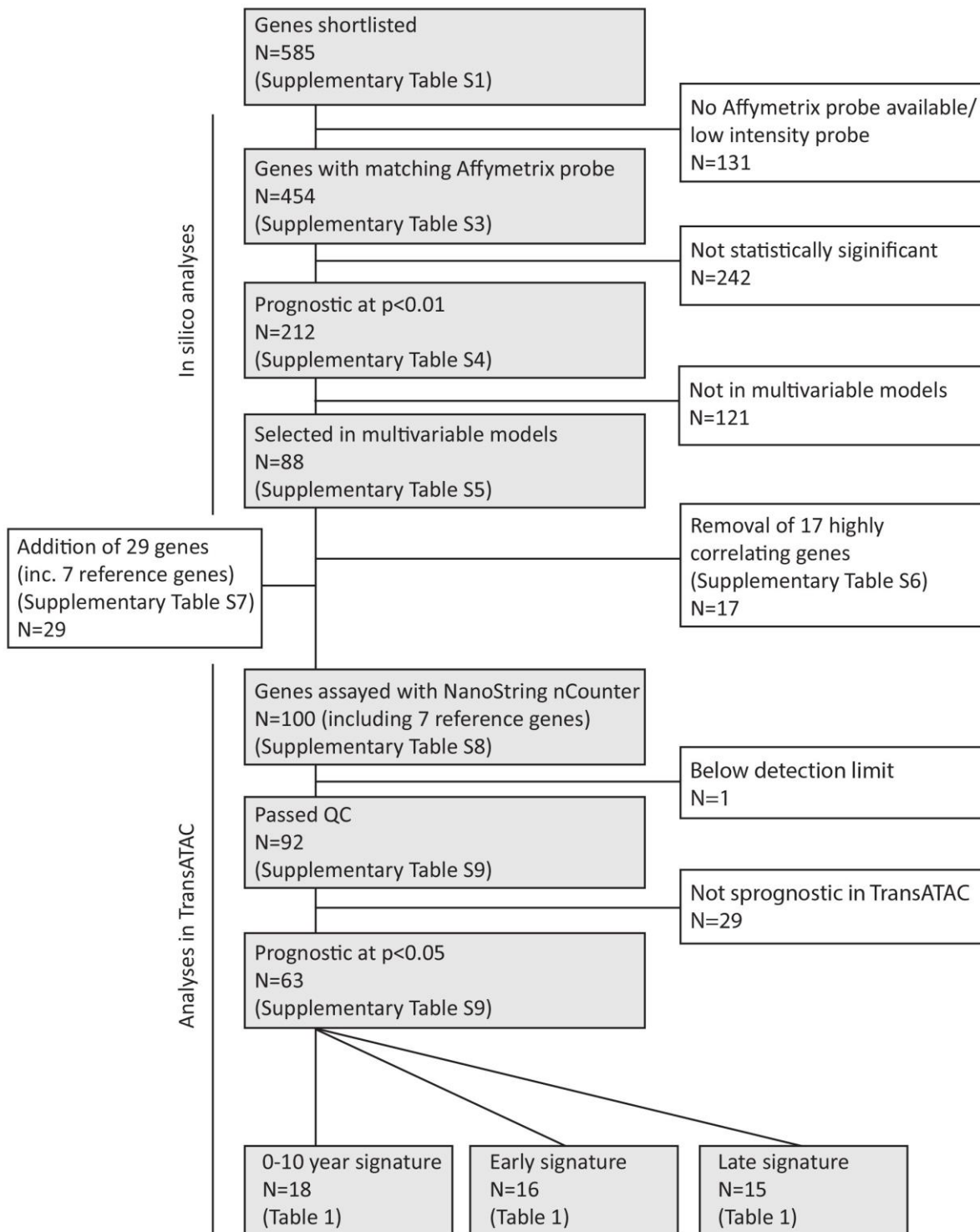


Figure 1. Flowchart of gene signature derivation *in silico* and in TransATAC cohort.

QC = Quality control; ATAC= Arimidex, Tamoxifen, Alone or in Combination.

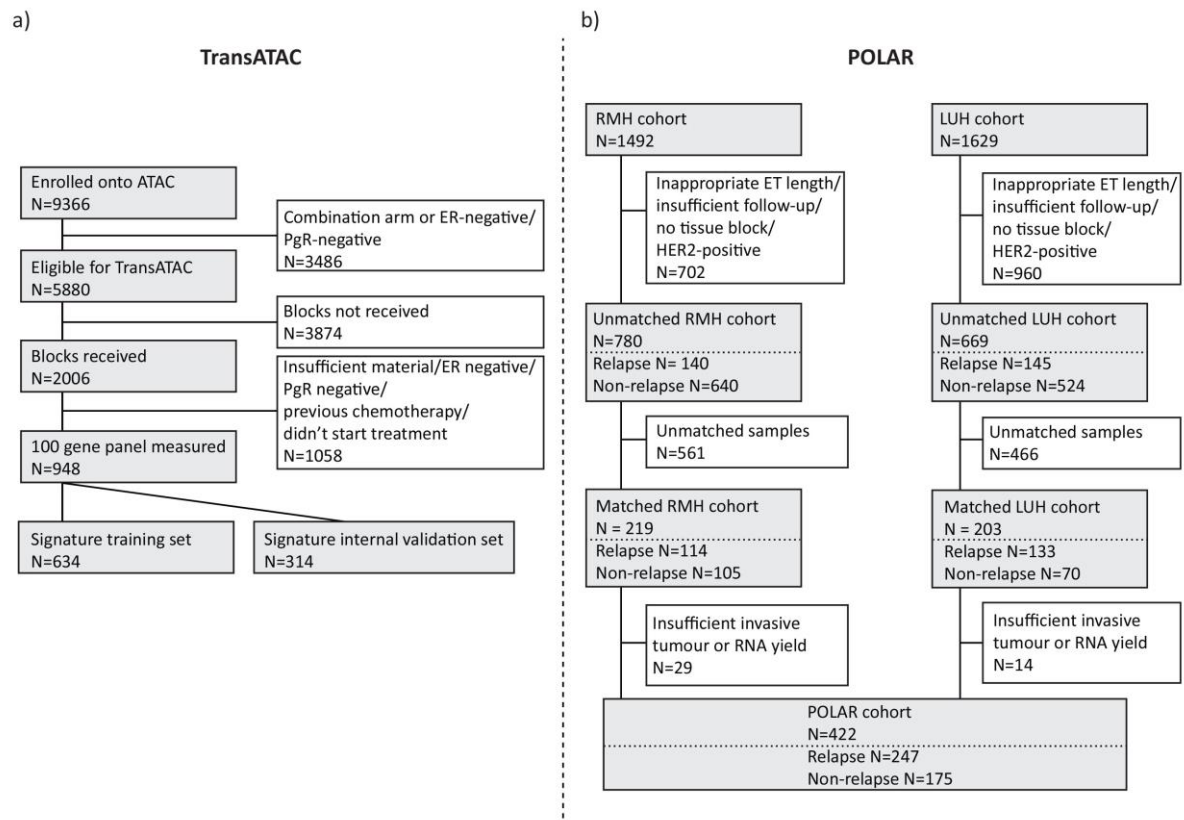


Figure 2. CONSORT diagram of the availability of samples for analysis from the (a) ATAC trial and (b) POLAR collection of samples.

Variable	10-year signature	Early signature	Late signature
ALDH1A1			-0.194
ANP32E	0.143	0.010	0.083
CRABP2	0.084	0.207	
CXCL12		-0.183	
CXCR4	0.142		0.056
EGFR			-0.030
ELF5	-0.046	-0.001	
FGF2	-0.178		-0.232
IGF1	-0.029	-0.017	
IGJ	-0.086	-0.037	-0.030
IL6ST		-0.044	
LINC00341	-0.463	-0.362	-0.392
LRRC48		-0.104	
MMP9	0.043		0.064
MPZL1	0.276	0.066	0.043
NUSAP1	0.088	0.065	
PBX1	0.159		0.375
PDZK1	-0.011		-0.063
PGR		-0.073	
PRC1		0.019	
RGL1	-0.429	-0.166	-0.161
RRM2	0.077	0.124	
SFRP1	-0.017		-0.278
STC2	-0.087	-0.068	
TNF	-0.029		-0.026
ZEB2			-0.138
CTS	0.514	0.409	0.516

Table 1. Variables and corresponding beta-coefficients of the time-dependent 10-year, early and late signatures.

Score	No. of patients (relapses)	Univariate comparisons					Multivariate comparisons: CTS+signature vs CTS				
		LR χ^2	p	HR (95% CI)	c- inde x	P diff	$\Delta L R \chi^2$	p	HR (95% CI)	c- inde x	P diff
0-10 years											
CTS	314 (59)	48.7	<0.001	2.16 (1.79 - 2.62)	0.674	-	-	-	-	-	-
10-year signature		28	<0.001	1.98 (1.54 - 2.55)	0.671	reference	7.9	0.005	1.49 (1.13 - 1.96)	0.709	reference
Early signature		33.4	<0.001	2.06 (1.62 - 2.61)	0.678	0.334	10.3	0.001	1.55 (1.19 - 2.02)	0.711	0.48
Late signature		18.1	<0.001	1.72 (1.34 - 2.20)	0.642	0.000	4.3	0.037	1.33 (1.02 - 1.74)	0.7	0.004
0-5 years											
CTS	314 (26)	29	<0.001	2.04 (1.53 - 2.74)	?	-	-	-	-	-	-
10-year signature		14.1	<0.001	2.05 (1.41 - 2.98)	0.678	0.833	3.2	0.073	1.46 (0.97 - 2.19)	0.712	0.77
Early signature		14.9	<0.001	2.00 (1.42 - 2.81)	0.672	reference	2.8	0.096	1.40 (0.95 - 2.06)	0.705	reference
Late signature		8.9	0.003	1.77 (1.22 - 2.57)	0.648	0.138	1.7	0.109	1.31 (0.87 - 1.97)	0.705	0.65
5-10 years											
CTS	270 (33)	19.8	<0.001	1.84 (1.33 - 2.54)	?	-	-	-	-	-	-
10-year signature		13.9	<0.001	1.93 (1.37 - 2.72)	0.663	0.027	4.8	0.028	1.53 (1.05 - 2.22)	0.696	0.14
Early signature		18.6	<0.001	2.11 (1.52 - 2.94)	0.681	0.091	8	0.005	1.70 (1.19 - 2.43)	0.708	0.14
Late signature		9.3	0.002	1.68 (1.21 - 2.34)	0.636	reference	2.7	0.099	1.36 (0.95 - 1.94)	0.686	reference

Table 2. HRs (95% CI), LR χ^2 , p-values and c-indices for CTS, 10-year, early and late signatures in TransATAC validation cohort.

Both univariate and multivariate analyses are presented for years 0 to 10, years 0 to 5, and years 5 to 10 separately. Likelihood ratio test based on Cox proportional hazard models for univariate and multivariate analyses. Differences in likelihood ratio values ($\Delta LR\chi^2$) were used. CTS = clinical treatment score; LR = likelihood ratio; HR = hazard ratio; CI = confidence interval.

POLAR patient group		Prognostic signature	Significance measure	0-10 years	0-5 years	5-10 years
All POLAR patients	No. of patients (relapses)			n=422 (247)	n=308 (133)	n=285 (114)
		CTS	$LR\chi^2$	11.23	22.09	0.90
			p	0.0008	<0.0001	0.34
		10-year signature	$LR\chi^2$	12.44	14.59	4.18
			p	0.0004	0.0001	0.041
		CTS+10-year signature vs CTS	$\Delta LR\chi^2$	7.74	7.59	3.58
			p	0.0054	0.0059	0.0586
Chemotherapy-treated	No. of patients (relapses)			n=170 (99)	n=128 (57)	n=113 (42)
		CTS	$LR\chi^2$	7.75	15.25	0.11
			p	0.0054	0.0001	0.7437
		10-year signature	$LR\chi^2$	7.73	7.41	3.15
			p	0.0054	0.0065	0.08
		CTS+10-year signature vs CTS	$\Delta LR\chi^2$	6.71	5.95	3.12
			p	0.0096	0.0147	0.0773
Chemotherapy-naïve	No. of patients (relapses)			n=252 (148)	n=180 (76)	n=172 (72)
		CTS	$LR\chi^2$	6.1	10.08	2.19
			p	0.0135	0.0015	0.1387
		10-year signature	$LR\chi^2$	5.39	7.26	1.70
			p	0.0202	0.0071	0.19
		CTS+10-year signature vs CTS	$\Delta LR\chi^2$	2.47	2.72	0.82
			p	0.1158	0.0992	0.3647

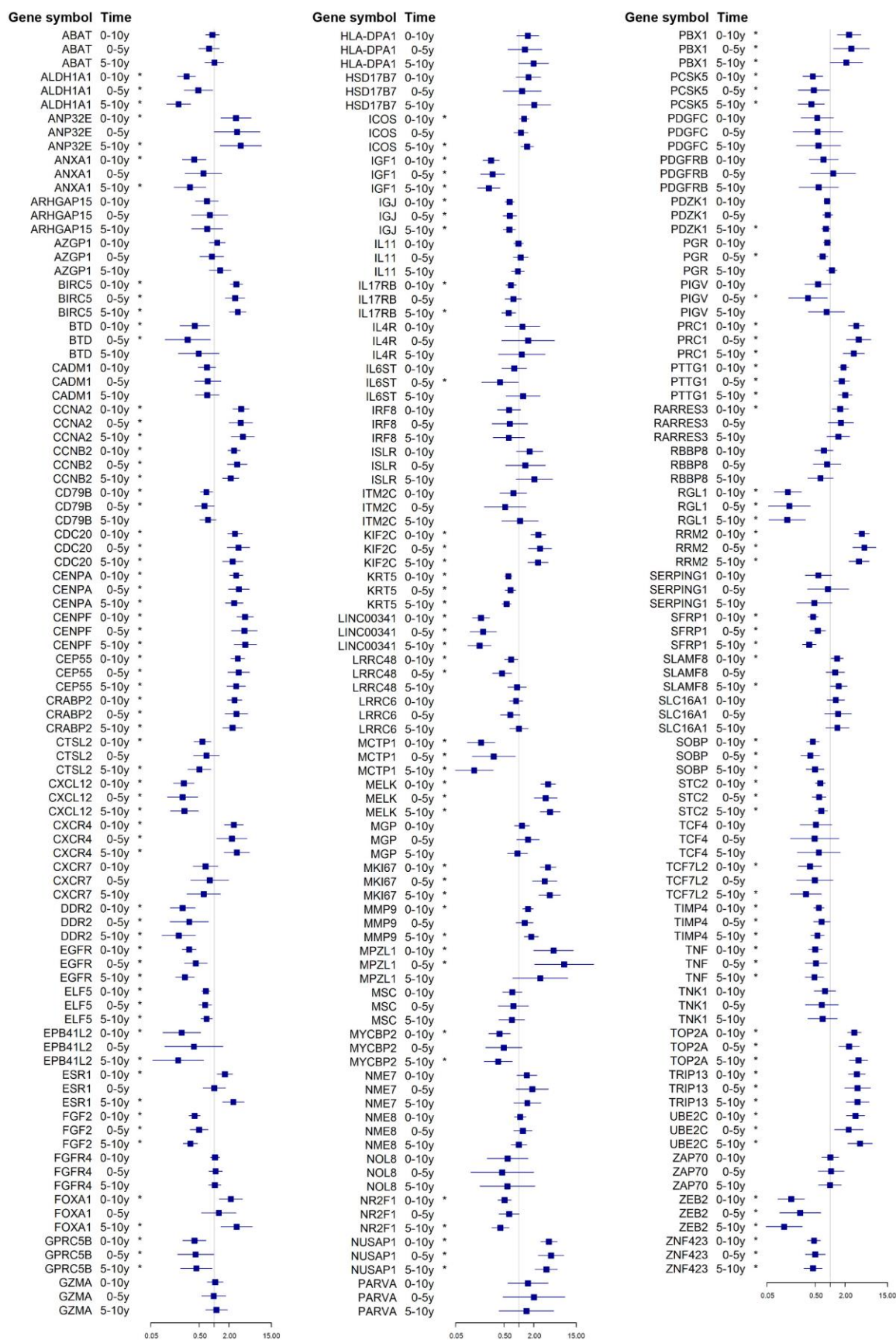
Table 3. $LR\chi^2$ and p-values for CTS and 10-year signature in three groups of POLAR validation set.

Both univariate and multivariate analyses are presented for years 0 to 10, years 0 to 5, and years 5 to 10 separately. Likelihood ratio test based on Cox proportional hazard models for univariate and multivariate analyses. Differences in likelihood ratio values ($\Delta LR\chi^2$) were used. POLAR = molecular Predictors Of early versus Late Recurrence in ER-positive breast cancer; CTS = clinical treatment score; LR = likelihood ratio.

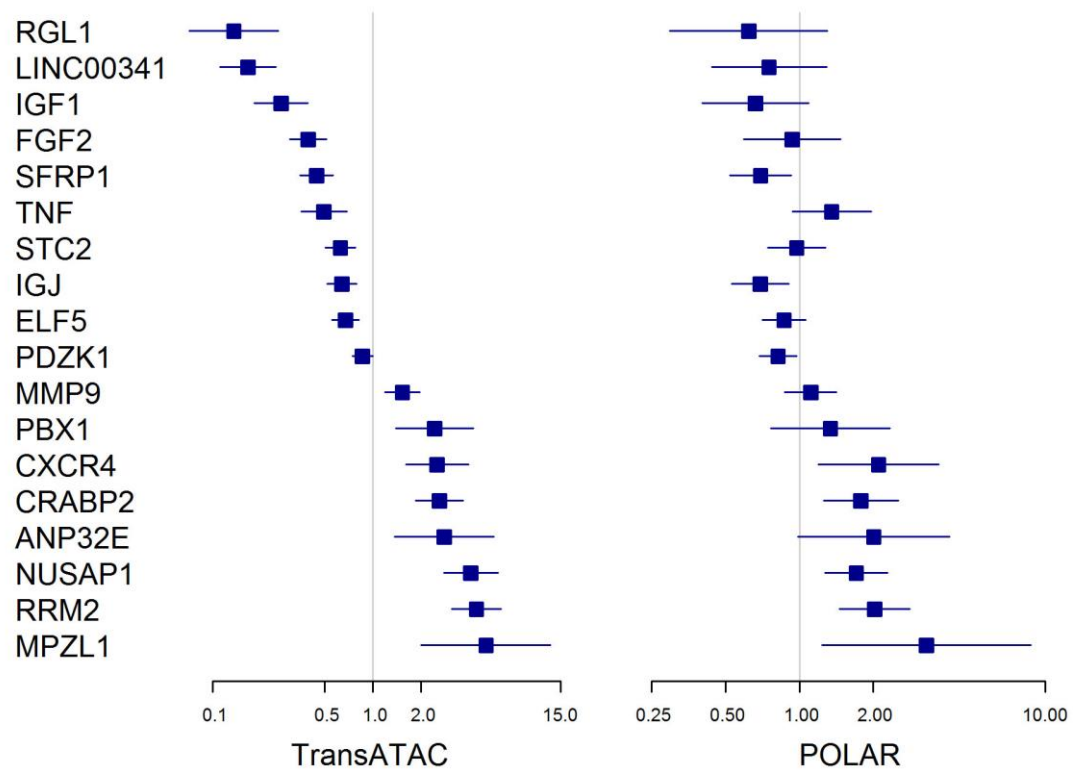
All recurrence (N=271, AR= 55)								
	Univariate				Multivariate comparison: CTS+signature vs CTS			
Score	LRχ^2	p	HR (95% CI)	c-index	ΔLRχ^2	p	HR (95% CI)	c-index
CTS	37.4	<0.001	1.94 (1.57 - 2.40)	0.664	-	-	-	-
10-year signature	20.7	<0.001	1.85 (1.42 - 2.41)	0.657	5.7	0.017	1.42 (1.07 - 1.89)	0.695
BCI	25.0	<0.001	2.07 (1.54 - 2.77)	0.679	9.5	0.002	1.62 (1.19 - 2.21)	0.711
RS	11.1	<0.001	1.52 (1.21 - 1.91)	0.607	5.8	0.016	1.35 (1.07 - 1.71)	0.683
ROR	18.3	<0.001	1.77 (1.36 - 2.31)	0.650	6.0	0.014	1.42 (1.07 - 1.87)	0.700
IHC4	14.4	<0.001	1.63 (1.28 - 2.10)	0.629	7.5	0.006	1.46 (1.12 - 1.91)	0.696
Distant recurrence (N=271, DR= 41)								
	Univariate				Multivariate comparison: CTS+signature vs CTS			
Score	LRχ^2	p	HR (95% CI)	c-index	LRχ^2	p	HR (95% CI)	c-index
CTS	46.7	<0.001	2.25 (1.79 - 2.82)	0.689	-	-	-	-
10-year signature	26.4	<0.001	2.24 (1.64 - 3.06)	0.694	8.5	0.004	1.65 (1.18 - 2.30)	0.733
BCI	34.0	<0.001	2.71 (1.91 - 3.84)	0.726	14.5	<0.001	2.03 (1.40 - 2.95)	0.754
RS	10.7	<0.001	1.58 (1.23 - 2.03)	0.616	5.1	0.024	1.38 (1.06 - 1.79)	0.707
ROR	21.3	<0.001	2.05 (1.50 - 2.79)	0.680	7.5	0.006	1.58 (1.14 - 2.21)	0.736
IHC4	17.9	<0.001	1.87 (1.41 - 2.49)	0.658	9.8	0.002	1.68 (1.22 - 2.31)	0.731

Table 4. HRs (95% CI), LR χ^2 , p-values and c-indices for CTS, 10-year signature, BCI, RS, ROR and IHC4 for all and distant recurrences respectively in the TransATAC validation cohort.

Both univariate and multivariate analyses are presented. Likelihood ratio test based on Cox proportional hazard models for univariate and multivariate analyses. Differences in likelihood ratio values (Δ LR χ^2) were used. AR = all recurrence; DR = distant recurrence; CTS = clinical treatment score; BCI = breast cancer index; RS = recurrence score; ROR = risk of recurrence; LR = likelihood ratio; HR = hazard ratio; CI = confidence interval.



Supplementary Figure 1. Forest plot of hazard ratios and confidence intervals for the 92 genes assessed in TransATAC in univariate analyses. * denotes significance.



Supplementary Figure 2. Hazard ratios and odds ratios for the 10-year signature genes in TransATAC and POLAR respectively

	TransATAC		
	All	Training	Validation
	n=948	n=634	n=314
Age at diagnosis*			
mean, years	64.5	64.1	65.2
median, years	63.7	63.6	64.2
Tumour size			
≤2cm	634	427	207
2-5cm	295	194	101
>5cm	19	13	6
Recurrence (all)			
	183	123	59
Treatment			
Radiotherapy	656	430	226
Mastectomy	373	262	111
Grade			
1	265	177	88
2	537	368	169
3	146	89	57
Nodal status			
node negative	665	441	224
1-3 positive nodes	199	136	63
4 or more nodes	84	57	27
Endocrine Rx			
Tamoxifen	464	301	163
AI	484	333	151

Supplementary Table S10. Demographics of TransATAC cohort

	RMH			LUH			POLAR (RMH+LUH)		
	Case	Control	Total	Case	Control	Total	Case	Control	Total
	114	105	219	133	70	203	247	175	422
Age at diagnosis*									
mean, years	56	58	57	62	60	61	59	58	59
median, years	54	58	56	61	58	61	58	58	58
range, years	29 - 93	28 - 88	28 - 93	35 - 100	35 - 87	35 - 100	29 - 100	28 - 88	28 - 100
Tumour size									
<2cm	40 (35%)	43 (41%)	83 (38%)	40 (30%)	27 (39%)	67 (33%)	80 (32%)	70 (40%)	150 (36%)
2-5cm	63 (55%)	54 (51%)	117 (53%)	89 (67%)	40 (57%)	129 (64%)	152 (62%)	94 (54%)	246 (58%)
>5cm	11 (10%)	8 (8%)	19 (9%)	4 (3%)	3 (4%)	7 (3%)	15 (6%)	11 (6%)	26 (6%)
Grade									
1	12 (11%)	14 (13%)	26 (12%)	9 (7%)	11 (16%)	20 (10%)	21 (9%)	25 (14%)	46 (11%)
2	48 (42%)	52 (50%)	100 (46%)	72 (54%)	35 (50%)	107 (53%)	120 (49%)	87 (50%)	207 (49%)
3	54 (47%)	39 (37%)	93 (42%)	52 (39%)	24 (34%)	76 (37%)	106 (43%)	63 (36%)	169 (40%)
Histological subtype									
IDC	80 (70%)	75 (71%)	155 (71%)	104 (78%)	55 (79%)	159 (78%)	184 (74%)	130 (74%)	314 (74%)
ILC	22 (19%)	18 (17%)	40 (18%)	27 (20%)	15 (21%)	42 (21%)	49 (20%)	33 (19%)	82 (19%)
other	12 (11%)	12 (11%)	24 (11%)	2 (2%)	0 (0%)	2 (1%)	14 (6%)	12 (7%)	26 (6%)
Nodal status									
node negative	60 (53%)	56 (53%)	116 (53%)	42 (32%)	29 (41%)	71 (35%)	102 (41%)	85 (49%)	187 (44%)
1-3 positive nodes	25 (22%)	34 (32%)	59 (27%)	54 (41%)	31 (44%)	85 (42%)	79 (32%)	65 (37%)	144 (34%)
4 or more nodes	29 (25%)	15 (14%)	44 (20%)	37 (28%)	10 (14%)	47 (23%)	66 (27%)	25 (14%)	91 (22%)
PgR									
Negative	5 (4%)	7 (7%)	12 (5%)	25 (19%)	9 (13%)	34 (17%)	30 (%)	16 (%)	46 (11%)
Positive	20 (18%)	23 (22%)	43 (20%)	102 (77%)	56 (80%)	158 (78%)	122 (%)	79 (%)	201 (48%)
Unknown	89 (78%)	75 (71%)	164 (75%)	6 (5%)	5 (7%)	11 (5%)	95 (%)	80 (%)	175 (41%)
NPI category*									
<3.4	25 (22%)	26 (25%)	51 (23%)	15 (11%)	13 (19%)	28 (14%)	40 (16%)	39 (22%)	79 (19%)
3.4-5.4	49 (43%)	50 (48%)	99 (45%)	77 (58%)	42 (60%)	119 (59%)	126 (51%)	92 (53%)	218 (52%)
>5.4	40 (35%)	29 (28%)	69 (32%)	41 (31%)	15 (21%)	56 (28%)	81 (33%)	44 (25%)	125 (30%)
Endocrine									

Therapy*									
Tamoxifen only	80 (70%)	72 (69%)	152 (69%)	96 (72%)	50 (71%)	146 (72%)	176 (71%)	122 (70%)	298 (71%)
Any AI	34 (30%)	33 (31%)	67 (31%)	37 (28%)	20 (29%)	57 (28%)	71 (29%)	53 (30%)	124 (29%)
Chemotherapy*									
No	54 (47%)	49 (47%)	103 (47%)	94 (71%)	55 (79%)	149 (73%)	148 (60%)	104 (59%)	252 (60%)
Yes	60 (53%)	56 (53%)	116 (53%)	39 (29%)	15 (21%)	54 (27%)	99 (40%)	71 (41%)	170 (40%)

Supplementary Table S11. Demographics of POLAR cohort

Supplementary methods

Patient cohorts

In silico cohort: 102 duplicate samples that were originally published in GSE6532 and subsequently reanalysed under GSE17705 were included in our cohort only once. The ER status of samples in GSE6532, GSE9195 and GSE17705 cohorts were defined in the original publications by immunohistochemistry. For GSE26971, probe intensity of 205225_at (ESR1) was used with the cut-off of 1000 also used in the original publication resulting in the removal of 7 samples from GSE26971. The HER2 status in all four cohorts was defined by using HER2 probe intensity (216836_s_at) with the cut-off of 6000. Forty-four samples were found to be HER2-positive by this criterion and subsequently removed from the cohort. In the 0-10 year follow-up period 741 samples had 168 DMFS events; the 318 samples with RFS data available had 83 relapses recorded. Nodal status was available for 696 samples.

Analytic procedures

For POLAR archival formalin fixed paraffin-embedded (FFPE) tissue blocks from either surgical excision specimens or core biopsies were identified. For patients who had been treated with neoadjuvant therapy, the diagnostic core was used for analysis. Oestrogen-receptor (ER), progesterone receptor (PgR) and HER2 status were determined from histopathology reports at diagnosis. For patients in whom HER2 status was unknown at the original diagnosis, HER2 staining was performed initially by immunohistochemistry (IHC) (graded from 0 to 3+) with confirmation of HER2 2+ tumours by in situ-hybridization (D-DISH). If tumours were subsequently identified as HER2-positive, they were excluded from the cohort.

Statistical methods

For the *in silico* study raw expression data of 22,277 probes was MAS5.0 normalized on a platform basis using the justMAS function in the simpleaffy R library to a mean target intensity of 600, without background correction. The two normalised datasets were then merged. Cross-cohort batch

effects were corrected using the COMBAT (sva R package, Surrogate Variable Analysis) empirical Bayes method (Johnson et al, 2007), directly removing known batch effects. Expression probes that had a <10% intensity of all probes were removed. Seventy-five genes had no corresponding probes in the assay, 510 genes had 933 probes associated in the assay, of which for each gene the highest variance was selected. Cox Proportional Hazard was used with both continuous and median split expression to identify significantly prognostic genes. Hazard ratios and Odds ratios were derived from the standard deviation of the Cox-model regression coefficient. Analyses were performed in the early, late and 10 year time periods with distant metastasis-free survival (DMFS) and relapse-free survival (RFS) as endpoints respectively. Statistically significant genes in univariate analyses were entered into multivariate Cox proportional hazards models.

In TransATAC the 92 genes were evaluated in the 948 sample set by continuous univariate Cox Proportional Hazard, genes significant at $p < 0.05$ were taken forward for signature generation. For this, the 948 patients were randomly split into 2/3 training ($n=634$) and 1/3 validation ($n=314$) sets. Number of events were split similarly, nodal status and tumour size were matched between training and validation sets. Genes statistically significant in univariate analyses in either early, late or 10-year periods and clinical treatment score (CTS) were entered into multivariate selection process. Elastic net penalised Cox regression was used for feature selection with leave one out cross-validation. The minimum partial likelihood deviance was estimated for different alpha values. This was done by varying the lambda tuning parameter, that controls the overall level of shrinkage. Leave-one-out cross-validation of the partial likelihood deviance was used to estimate the best lambda. The partial likelihood deviance given alpha and lambda was obtained by getting a 1 model to all data except one observation. Then the deviance difference was between all data, and a calculation not using the observation left out. This was repeated for all data points. The lambda for each alpha was chosen was based on a 'one-standard-error' rule. This selects the model with deviance one standard error away from the minimum. Alpha was set at 0.2 for all three model selections. Composite scores were built using the selected features and the beta-coefficients were determined. Beta-coefficients were normalised by dividing the gene standard deviation of the training population.

In POLAR the primary analyses were performed on all POLAR patients for which NanoString data was available and passed QC criteria. A control was defined as a patient who did not relapse during follow up. Controls were randomly selected according to matching criteria from the remaining cohort of patients who did not relapse during follow-up. The four matching criteria used in this study were: (i) age at diagnosis (≤ 50 years, > 50 years), (ii) Nottingham Prognostic Index (NPI) category (< 3.4 ; $3.4-5.4$, > 5.4), (iii) type of adjuvant endocrine therapy (tamoxifen only, any aromatase inhibitor), (iv) chemotherapy use (yes, no). All parametric unpaired t-tests were performed using PRISM software (Version 6.0c). Conditional logistic regression was performed using STATA to test whether the 10-year signature, CTS and individual genes were associated with risk of recurrence in a non-pairwise fashion. A multivariate conditional logistic regression analysis was performed to see whether CTS and the 10-year signature were independent variables in a forward selection manner with a 5% significance level in a non-pairwise fashion. A log likelihood test was used to test the 10-year signature and CTS score in a model provides a better fit than CTS alone with a 5% significance level. The two separate hospital cohorts constituting POLAR were also analysed separately, no significant difference in the results was found (data not shown).

Reference

Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007;8(1):118-27.